A Proximity based Stress Testing Framework

Boris Waelchli
Department of Banking and Finance, University of Zurich,Plattenstrasse 14
CH-8032 Zuerich, Switzerland
boris.waelchli@uzh.ch

October 12, 2015

Abstract

In this a paper a non-linear macro-stress testing methodology with focus on early warning is developed. The methodology builds on a variant of Random Forests and its proximity measures. It develops a framework in which naturally defined contagion and feedback effects transfer the impact of stressing a relatively small part of the observations on the whole dataset and thus allow to estimate the a stressed future state. It will be shown that contagion can be directly derived from the proximities while iterating the proximity based contagion leads to naturally defined feedback effects. This procedure allows accurate forecast of events under stress and thus allows to forecast the emergence of a potential crisis. The framework also estimates a set of the most influential economic indicators leading to the potential crisis, which can then be used as indications of remediation or prevention of that crisis. Since the methodology is Random Forests based the framework is suitable for big data analysis.

Keywords: Random Forests; Recursive Conditional Partitioning; Stress Testing; Early Warning Indicators; Big Data

Contents

List of Tables

1 Introduction

Stress testing is of increasing importance in all industries. Regulatory requirements as well as renewed accounting standards are asking for macro stress tests to better safeguard against a crisis. Macro stress testing is a relatively new field though. Its merits are seen in the context of either crisis management or early warning indication. To manage a crisis a stress scenario is applied to known key risk indicators (KRI) and, for example, the resulting economic capital is determined whereas in an early warning indication framework the KRIs are identified. In the case of early warning, scenario design is crucial (Borio et al. 2012). Ideally, the macro prudential scenarios to be used should be plausible, severe and suggestive of mitigation opportunities (Breuer and Krenn 1999). For example in the shadow of the ongoing crisis scholars (Borio et al. 2012) are suggesting that scenarios might have to be implausibly severe to include the expectation of the unexpected. Apart from the obvious choice of historical scenarios, measures for plausibility of self constructed, hypothetical scenarios and algorithms and methods to find them have been suggested. Especially in the last case scholars have formulated doubt as to whether early warning frameworks can actually work (Borio et al. 2012). The failure of prediction ahead of the financial crisis in 2007/08 indeed casts doubt on the usage of historical data to assess the probability of an upcoming crisis.

From a methodological point of view, researchers find that most currently performed macro stress tests do not go beyond the first year effects and could be enhanced by a longer time horizon and corresponding correlation/contagion and feedback effects, preferably in a nonlinear framework (Borio et al. 2012). Additionally it is often assumed that modelled interdependence remains stable over time while in stressed situations relations can change quickly (Borio et al. 2012).

At the BIS it is propagated that macro stress testing is a toolbox, not a single tool. This paper adds a tool by developing a framework that is a big data suitable model for adverse economic movements with high predictive capabilities, taking only a few inputs with the option for finding policy indicators. The proposed model is based on the proximities of a Random Forest variant on an empirical dataset. The framework now does not focus on variables and risk indicators but on a subset of observations. More specifically a suitable subset of observations from an initial dataset today are chosen and stressed on all variables of each of this observations. This stress should reflect the values of the variables of the stressed observation in future state, seen from the initial dataset. How this observations are then stressed is not the subject of this paper, it can be done taking into account common econometrical models of interaction between the variables or expert judgement. As pointed out, the proposed framework does not cover this step. Once the chosen sample is stressed, by means of the Random Forest proximities all other, not sampled, observations are infected by the values of the variables of the stressed observations. This step is often called contagion. On the infected, estimated future dataset a new Random Forest model is built. This paper shows that the resulting Random Forests model predicts future stress events accurately using a relatively small initial sample of observations. The model encompasses the concept of contagion and can be iterated to produce feedback effects, inherently defined in the model. Based on the estimated number of stress caused events in the future dataset, the framework indicates the whether a potential crisis could emerge due to the applied scenario. In case the framework suggest a crisis, the importance measures defined within the Random Forests algorithm allow to identify the most important variables which had the highest influence in the classification of the observations. This variables can then be used to identify remediations against the potential crisis. Thus the proposed proximity based stress testing framework can be used as an early warning indicator as well as instrument to identify actions to manage or prevent a crisis.

Being based on Random Forest the methodology can cope with a large number of indicators being thus suitable for big data analysis and can consequently model national and international KRIs together.

The remainder of the paper is structured as follows: the second section positions the paper in the current literature on macro stress testing and early warning frameworks under special consideration of Random Forest. In the following third section the concept of the Random Forest variant of recursive conditional participation and the proximities is introduced. The fourth section lays out the mathematical foundation of the proposed model, the following section applies the model to an empirical analysis while the sixth section elaborates the policy indications from the model. The final section concludes the paper.

2 Literature Review

The contribution of this paper is situated in three areas of the macro stress testing literature: in general macro stress test modelling, in modelling early warning frameworks and in the application of machine learning algorithms.

Current trends in macro stress testing encompass integrating different risks, contagion- and feedback effects:
The idea of contagion is the transmission of a shock by a relatively small number of market participants to other or most of the other participants. To include the concept of contagion has become a common feature in macro stress testing. Some of the earlier works are by Allen and Gale (2000), who model contagion effects of claims between banks; De Bandt and Hartmann (2001), consider contagion effect in the broader context of systemic risk and Upper and Worms (2004) specifically analyse contagion in the german interbank market.
Feedback effects describe the effects of stress and also contagion spreading between the market participants in the subsequent periods of time after the shock and contagion have occurred. Feedbacks are for example modelled by Jacobsen et al. (2005) who use an aggregate vector autoregressive model integrating several modules linking risk factors and balance sheets of corporates to show feedback from financial stability to the economy.
Elsinger et al. (2006) integrate market risk, credit risk, interest rate risk and counterparty credit risk in the Austrian interbank sector. Boss et al. (2008) have extended the model of Elsinger et al to a 3 year horizon and incorporated profit risk. Considering the rules of accounting Drehmann et al. (2010) have created a stress test integrating credit and interest rate risk by modelling assets and liabilities simultaneously. State of the art stress tests also increasingly try to include liquidity risk (Barnhill and Schumacher 2011), (Van den End 2008). The currently most comprehensive model is the risk assessment model of the Bank of England (Aikman et al. 2009) which also includes feedback effects.

The BIS has in its various published analyses assessed that the focus on non-linearities and feedback effects is a priority while they doubt the potential of modelling network effects or aggregation models (Borio et al. 2012).
In their studies, Juselius and Kim (2011) and also Drehmann et al. (2007), have found that the macro econometric relationships are mostly non-linear. Again Drehmann and Tarashev (2011) and Borio and Drehmann (2009) find that crises often begin at the peak of a financial cycle.
The proposed model is non-linear, incorporates contagion and feedback effects and it will be shown that the performed stress tests ahead of the crisis are accurate. On the other hand, the model is only integrated in the sense that the dependent variable depends on various macroeconomic factors from different areas of risk but they do only influence each other indirectly by changes in proximities. However, since the proposed model does not specify how the stressed observations are built, another model from the literature which integrates all risks types can be applied to generate the stressed sample of observations which is used in the proposed framework.

Generally the usage of stress testing for early warning indication is not recommended by BIS (Borio et al. 2012). Reasons are often the lack of non-linearity and the usage of historical scenarios. The proposed framework is focused on early warning yet it is not built in the classic way using early warning indicators: The early warning literature in finance mainly encompasses two approaches, signalling approaches, where a threshold for specific indicators is identified and logit/probit models modelling again the effects of early warning indicators. The proposed framework on the other hand estimates the number of events under stress modelling the interaction of a large of even vast amount of indicators and then reverts back to identify the most important of the indicators for remediation of the stress effects. Additionally it is inherently non-linear. However, the initial estimation is done on empirical data.

Regarding scope of data and indicators a recent representative of the signalling approach is Pasricha et al. (2013) who apply an imbalance indicator model encompassing a large number of potential indicators. While alternatively in a recent work Babecky et al. (2012) focus on developed economies and find by Bayesian model averaging that domestic housing prices, share prices, and credit growth, as well as the global variable private credit are KRIs.
The proposed model on the other hand encompasses around 100 indicators and mostly developed countries but also some emerging economies.

With focus on the proposed application of a Random Forest model several preceding papers can be identified. The first paper is Ghosh and Ghosh (2002) followed by Frankel and Wei (2004) who both apply decision trees on currency crises. Manasse and Roubini (2009) use binary recursive trees on sovereign crises. The succeeding paper by Savona and Vezzoli (2015) deal again with sovereign crises, while Duttagupta and Cashin (2011) and Manasse et al. (2013) study banking crises in emerging markets. Alessi and Detken (2014) apply regression trees to excessive credit growth and leverage measurement. Savona and Vezzoli (2015), Manasse et al. (2013) as well as Alessi and Detken (2014) all run some sort of Random Forest on their sample. Especially Alessi and Detken run the classic regression Random Forest by Breiman and Cutler (2005) to identify the most important variables to build the forest and construct a final decision tree with this important variables only.
This paper applies likewise a Random Forest model but firstly not by using the classic Random Forest by Breiman but the conditional recursive partitioning forest by Hothorn et al. (2006b)/Strobl et al. (2007) and not for the aim of building a final tree but to construct a stressed (future) dataset. The conditional recursive partitioning framework is chosen above the classic Random Forest because the latter is known to be biased in the choice of splitting variables and thus in the assessment of the most important variables. The classic Random Forest prefers continuos variables to factors or discrete variables or variables with many different values to such with less values on the observations. Since this paper also aims at identifying the most important variables the unbiased method is preferably chosen.
Additionally the proposed model is based on a classification forest and not a regression forest. This is done because the classification model can be stressed and interpreted in

an intuitive way: a stress situation will result in more observations being classified as events. On the other hand, regression trees and regression forest do not really produce new estimates of the dependent variable when new data is applied. Yet for stress testing this is precisely the aim.

3 Mathematical Background

3.1 Recursive Conditional Partitioning

The following description follows (Waelchli 2015).

The conditional recursive partitioning forest and the classic Random Forest are very similar. The main difference is the splitting framework: within the conditional recursive partitioning framework the variables for splitting are selected based on which has the highest association to the dependent variable based on a linear statistic. In the classic algorithm the homogeneity of a n dimensional space is optimized. Like the classic Random Forest algorithm, the knots in each tree are split on a random sample of the total variables. Unlike Random Forests, the trees are not grown on bootstrap samples but on samples without replacement. Strobl et al. (Strobl et al. 2007) have shown that the bootstrap samples increase the bias in variable selection identified in the classic Random Forests. The conditional recursive partitioning framework then grows each tree in the forest in accordance with the following rules (Hothorn et al. 2006b):

1. At each knot, test the global hypothesis of independence between the dependent variable and the drawn set of independent, potential splitting variables is tested. If the hypothesis cannot be rejected, and independence is assumed, the growth of the tree is stopped in the respective branch. If the hypothesis is rejected, in accordance with a predefined confidence level, the association of each dependent variable with the independent variable is tested and the variable with the highest association, as measured by the highest statistical significance (p value), is chosen as the variable to split on.

2. On the variable with the highest association, the point for the best binary split is the value which maximizes the test statistics for association. The data in the respective knot is split by that value alike in the classic Random Forest.

3. The steps are repeated within each tree for all trees in the forest until the global null hypothesis can not longer be rejected or another stopping criteria, for example, this means that the minimum number of observations in the respective knot applies.

To assure that the statistical tests in this framework do not require knowledge of the distribution of the underlying data, permutation testing is applied, where all possible permutations of the values in the learning sample are tested. For more details on the permutation tests, please refer to the framework of conditional inference and permutation tests (Strasser and Weber 1999). Thus alike within the classic Random Forest, it is not necessary to know or estimate the distributions of the variables or risk factors. (Strobl et al. 2007) have shown that he conditional inference framework is not biased in the choice of splitting variables. The analysis can thus be used to assess the most important variables. In this paper the permutation variable importance measurement is applied, which assesses the difference between the prediction of correctly forecast events before and after a respective variable has been randomly permuted. The implementation of the measure in the R package of cForest is used (Hothorn et al. 2006a).

For further details on the theory of the algorithm of the conditional inference framework, please refer to (Strasser and Weber 1999) and (Hothorn et al. 2006b). However the focus in this paper is on the proximity measure which thus will be defined in detail here.

Assuming a dataset with n independent variables $\mathbf{X} := (\mathbf{X}_1, ..., \mathbf{X}_n)$, with $\mathbf{X}_i := (X_{1i}, ..., X_{mi})$, one dependent variable $\mathbf{Y}$, with $\mathbf{Y} := (Y_1, ..., Y_m)$ and m observations $\mathcal{O}_j := (Y_j, \mathbf{X}_j)$ for $j \in \{1, ..., m\}$, while $\mathcal{O} := (\mathbf{Y}, \mathbf{X})$. For this purpose, $\mathbf{Y}$ is binary and describes the outcome event/non-event thus $Y_j \in \{0, 1\}$.

Definition 1. *For a dataset of observations $\mathcal{O} \in \mathbf{R}^{1xm}$ and a single observation $\mathcal{O}_{i.}, i \in \{1, ...m\}$ a Random Forest (prediction) is defined as*

$$RF(\mathcal{O}) \in \{0,1\}^{1xm}, \qquad RF(\mathcal{O}_{i.}) \in \{0,1\} \tag{1}$$

, assuming a binary (0,1) classification.

The proximity measurement is specifying a concept of distance between two observations in a Random Forest model. For each pair of observations, $\mathcal{O}_i$ and $\mathcal{O}_j$, their proximity ρ_{ij} is defined as the average number of final knots in all trees in the forest where both observations are in the same final knot.

Definition 2. *For two observations $\mathcal{O}_i$ and $\mathcal{O}_j$, $i, j \in \{1, .., m\}$, in a Random Forest with K trees, the proximity ρ_{ij} is defined as*

$$\rho_{ij} := \frac{1}{K} \sum_{k=1}^{K} \sum_{\tilde{k} \in \tilde{K}_k} I(\mathcal{O}_i \in \tilde{k}) I(\mathcal{O}_j \in \tilde{k}). \tag{2}$$

$\tilde{K}_k$ is the set of final knots in the k-th tree of the forest and $\tilde{k} := \{\mathcal{O}_{1\tilde{k}}, ..., \mathcal{O}_{m\tilde{k}}, i\tilde{k} \in \{1, ..., m\}\}$ is a single final knot within this set, thus $\tilde{k}$ is the set of observations which end up in this specific final knot after being applied to the classification tree. I is the indicator function.

The proximity measure has the following characteristic:

$$\rho_{jj} = 1 \;\; \forall j \in \{1, ..., m\}, \qquad 0 < \rho_{ij} < 1 \;\; \forall i \neq j \in \{1, ..., m\} \tag{3}$$

3.2 Mathematical Derivation of the Proximity based Stress Testing Framework

The idea of proximity based stress testing is simple: the closer two observations are, the more likely they influence each other. Additionally there is no need to stress all observations but only a preferably small share of the total number of observations and then let the contagion and feedback effects do the rest.
At the end, the aim is to take a dataset of interest, choose a specific number of observations, stress those based on expert judgement or on an econometric model and apply the proposed model to construct a stressed future dataset by contagion and feedback effects. On the stressed estimated dataset a new Random Forests model is built resulting in the estimated stressed state of all observations and a list of the most important variables leading to it.
In detail, in the constructed framework the contagion and feedback effects are done by proximity weighted averages of the stressed inputs or, for the purpose of modelling feedbacks, by proximity weighted averages of all other values on a specific variable respectively. A proximity weight is simply the relative proximity between two observations i and j:

Definition 3. *Definition of proximity weights ρ^{ω}:*

$$\rho^{\omega}_{ij} = \frac{\rho_{ij}}{\sum_{k=1}^{m} \rho_{ik}} \tag{4}$$

To derive the above described application of the framework, the following assumption must hold:

Assumption of Structural Stability - The proximities of observations develop similarly over time. Thus for two sets of data, which are sufficiently close in time, the proximity matrices are equal. In other words, for each pair of observations an $\epsilon > 0$ exists such that:

$$[\rho^{t}_{ij}] = \boldsymbol{\rho}^{t} = \boldsymbol{\rho}^{t+\epsilon} = [\rho^{t+\epsilon}_{ij}] \tag{5}$$

It can be shown that the dataset in a Random Forests prediction can be replaced by a dataset of proximity weighted averages. The latter can be shown to be generated by a subset of the observations and still yield the same predictions as the initial forest. It follows that, assuming structural stability, a generating subset for a future stress scenario can be used to build a fully stressed dataset which inserted in a current Random Forests model yields predictions of stressed events:

Proposition 1. *(Invariance of Prediction) The prediction results of a Random Forest model are equal for a specific dataset and the dataset of its respective proximity weighted averages: Assuming a perfectly accurate, fully grown Random Forests model with no sampling when growing the trees, and proximities larger than zero between the observations of each class, it follows that a positive integer n exists such that:*

$$RF(\mathcal{O}) = RF([\sum_{k=1}^{m} [\rho^{\omega}_{ij}]^{n}_{ik} \mathcal{O}_{kj}]_{ij}) \tag{6}$$

while RF(.) refers to the prediction of the forest of the events and non-events.

Proof. It needs to be shown that all observations are classified as the same class before and after the application of proximity based contagion: $\mathcal{O}_i \in \{\mathcal{O}_j : RF(\mathcal{O}_j) = cl\} \Rightarrow [\sum_{k=1}^{m} \rho_{ik}^{\omega} \mathcal{O}_{kj}]_{i.} := \hat{\mathcal{O}}_i \in \{\hat{\mathcal{O}}_j : RF(\hat{\mathcal{O}}_j) = cl\}, cl \in \{0, 1\}$

Because of the full growing of all trees, all nodes are homogenous and because of the usage of the full sample, there is no misclassification in any tree within the forest. Thus, training observations from different classes have a proximity measure of zero.

Without loss of generality it holds that the value $\mathcal{O}_{ij}$ of observation $\mathcal{O}_{i.}$ on variable $\mathcal{O}_{.j}$ is transformed, by proximity based contagion, into:

$$\hat{\mathcal{O}_{ij}} := \sum_{k \in \{k : \mathcal{O}_{k.} \in CL(RF(\mathcal{O}_{i.}))\}} \rho_{ik}^{\omega} \mathcal{O}_{kj} \tag{7}$$

where $CL(RF(\mathcal{O}_{i.})) := \{\mathcal{O}_{k.} : RF(\mathcal{O}_{k.}) = RF(\mathcal{O}_{i.})\}, RF(\mathcal{O}_{i.}) \in \{0, 1\}$. All transformed values of a variable $\mathcal{O}_{.j}$ are thus weighted averages of the values of observations in the respective same class only.

The applied matrix of proximity weights $[\rho_{ij}^{\omega}]$ can, without restriction to generality, be written as a block diagonal matrix, sorted by the classes of the dependent variable: In the rows and columns put first the observations with class one and second those with class zero:

$$[\rho_{ij}^{\omega}] = \begin{pmatrix} [\rho_{ij}^{\omega}]_{i,j:\mathcal{O}_{i.} \in CL(1), \mathcal{O}_{j.} \in CL(1)} & 0^{|CL(1)| * |CL(0)|} \\ 0^{|CL(0)| * |CL(1)|} & [\rho_{ij}^{\omega}]_{i,j:\mathcal{O}_{i.} \in CL(0), \mathcal{O}_{j.} \in CL(0)} \end{pmatrix} \tag{8}$$

Sorting the observations (rows) in the initial dataset $\mathcal{O}$ in the same way, the estimated dataset resulting from a one off application of proximity based contagion, $\hat{\mathcal{O}}$, can be written as matrix multiplication of $[\rho_{ij}^{\omega}]\mathcal{O}$. Iterating the estimated dataset means taking the proximity weighted average of the proximity weighted average iteratively. The proximity weighted average of the proximity weighted average is then $[\rho_{ij}^{\omega}][\rho_{ij}^{\omega}]\mathcal{O}$. This results in a power sequence:

$$[\rho_{ij}^{\omega}]^1 \mathcal{O}, [\rho_{ij}^{\omega}]^2 \mathcal{O}, ..., [\rho_{ij}^{\omega}]^n \mathcal{O}$$

$[\rho_{ij}^{\omega}]$ is per definition a stochastic row matrix as are its non zero diagonal blocks, as defined in equation 8. Within the non zero diagonal blocks, $[\rho_{ij}^{\omega}]_{i,j:\mathcal{O}_{i.} \in CL(cl), \mathcal{O}_{j.} \in CL(cl)}, cl \in \{0, 1\}$ the diagonal is non zero and always holds the highest row and column value (because each observation is closest to itself). Additionally it was assumed that observations within the same class have a non zero proximity.

Thus $[\rho_{ij}^{\omega}]$ has symmetrical square matrices on its diagonal, with no zero entries in the square matrices and no rows with only zero entries in the whole matrix. Having no entries in the upper right corner it is also of lower triangular form. Fulfilling this conditions Qu, Wang and Hull (Qu et al. 2005) have shown that the sequence of stochastic matrices of proximity weights, $[\rho_{ij}^{\omega}]^k$ converges:

$$lim_{k\to\infty}[\rho^{\omega}_{ij}]^k = \begin{pmatrix} 1_{|CL(1)|}c_1 & 0^{|CL(1)|*|CL(0)|} \\ 0^{|CL(0)|*|CL(1)|} & 1_{|CL(0)|}c_0 \end{pmatrix} \tag{9}$$

with c_1 and c_0 being stochastic vectors.

Assuming without restriction of generality that for each variable $\mathcal{O}_{.v}$, larger values of variable $\mathcal{O}_{.v}$ are associated with class 1 and lower with class 0 then the limits of each class are different. Then, because of the convergence of the sequence, it must hold that for each variable $\mathcal{O}_{.v}$ a positive number n_v of iterations exists, such that all proximities between class one observations at this point in the sequence, are larger than all proximities between class zero entries:

$$\exists n_v : min([\rho^{\omega}_{ij}]^{n_v}_{i,j:\mathcal{O}_{i.}\in CL(1),\mathcal{O}_{j.}\in CL(1)}\mathcal{O}_{i,v:\mathcal{O}_{i.}\in CL(1)}) >$$
$$max([\rho^{\omega}_{ij}]^{n_v}_{i,j:\mathcal{O}_{i.}\in CL(0),\mathcal{O}_{j.}\in CL(0)}\mathcal{O}_{i,v:\mathcal{O}_{i.}\in CL(0)})$$

Choosing the number of iterations n as $n := max_v n_v$ and because $\mathcal{O}$ can be perfectly predicted by a random rorests model, then $\rho^{\omega}_{ij}]^{n_v}\mathcal{O}$ can likewise be perfectly predicted by a random forest model and the observations are classified in the same class as by the initial forest.

□

Corollary 1. *(Minimal generator) For each dataset $\mathcal{O}$, a minimal generating set in $\mathcal{O}$ exists which, by application of proximity weighted averages generates a dataset $\hat{\mathcal{O}}$ such that the same Random Forest predictions result as would for RF($\mathcal{O}$): Assuming a perfectly accurate, fully grown Random Forests model with no sampling when growing the trees, and proximities larger than zero between the observations of each class, it follows that a minimal generating set $\mathcal{S} \subset \mathcal{O}$ exists which generates the same RF results as RF($\mathcal{O}$) using proximity based contagion:*

$$RF(\hat{\mathcal{O}}) = RF([\sum_{k:\mathcal{S}_{k.}\in\mathcal{S}}[\rho^{\omega}_{ij}]^n_{ik}\mathcal{S}_{kj}]_{ij}) = RF([\sum_{k:\mathcal{O}_{k.}\in\mathcal{O}}[\rho^{\omega}_{ij}]^n_{ik}\mathcal{O}_{kj}]_{ij}) = RF(\mathcal{O}) \tag{10}$$

Proof. Following the proof of proposition 1 every row $i : \mathcal{O}_{i.} \in \mathcal{O} \setminus \mathcal{S}$ is set to a vector of zeros. Lets call this dataset $\mathcal{O}_{\mathcal{S}}$. The matrix of proximities $[\rho^{\omega}_{ij}]^k$ is not changed as it is built by $RF(\mathcal{O})$.
The resulting proximity weighted average for a chosen value is built from sampled observations within the same class based on the proximities between the sampled observations and the observation from which the estimated value stems. Obviously the changes in the initial dataset do not affect the convergence of the stochastic proximity weight matrix (Qu et al. 2005).
Because the random forest model is again assumed to the perfectly accurate it holds again that on each variable $\mathcal{O}_{.v}$ the higher values can be attributed to a certain while the lower ones, after a certain threshold, can be attributed to the other class. Let's attribute again, without loss of generality, the higher value on each variable to class 1.

If $\mathcal{S} \subset \mathcal{O}$ but $\mathcal{S} \not\subset \mathcal{O} \setminus CL(1) \wedge \mathcal{S} \not\subset \mathcal{O} \setminus CL(0)$, thus $\mathcal{S}$ contains observations of both classes. Then the multiplication of the power sequence in equation 9 and $\mathcal{O}_{\mathcal{S}}$ again converges to a dataset $\hat{\mathcal{O}}$ whit the same Random Forests predictions as the inital dataset $\mathcal{O}$.

If on the other hand $\mathcal{S} \subset \mathcal{O} \setminus CL(j)$, $j \in \{0, 1\}$ then the observations of one of the two classes in $\hat{\mathcal{O}}$ are zero. However, since the values on the other class are larger than zero a random forests model can discriminate them again perfectly.

As such the assumed perfect accuracy of the forest is preserved by the sampled transformation and the prediction remains as on the initial forest using the whole dataset.

$\square$

Remark 1.

- *The assumption of non-zero proximities within the diagonal stochastic square matrices reflects the expectation that if a model is built and accurate, the observations of the same class are sensitive to the same risk drivers in many if not most cases. However, in the empirical application there will be cases where this assumption does not hold.*

- *Additionally the assumption of a perfectly accurate forst will not always hold on empirical dataset. However, using a large amount of variables and a full training sample, experience has shown that the forests models are almost perfectly accurate.*

- *The forest is built on the full dataset and not on training samples as is usually done (Hothorn et al. 2006a). Due to this assumption the stochastic matrix of proximity weights convergences to a block diagonal matrix made of two distinct stochastic row vectors. In case the data is sampled and the proximity matrix is non-zero, the stochastic matrix of proximity weights would converge to a matrix made of only one row vector and the resulting proximity transformed values would be equal regardless of their class. On the other hand refraining from sampling makes the forest prone to overfitting. However, this condition is uncommon in the literature. Also it is likely that proximities of observations of different classes tend to be low or zero if the forest is sufficiently accurate. In this cases the assumption would not be necessary. In case the proximities of observations in different classes are higher than zero but very low, it can be stated that the first step of the iteration has the largest effect on the data and the convergence will only occur late and slow. Thus in the practical application a sampling can be assumed while the forests should be calibrated to be as exact as possible.*

Corollary 2. *(Stress Prediction) Assuming a perfectly accurate, fully grown Random Forests model with no sampling when growing the trees, proximities larger than zero between the observations of each class, and a time series of datasets of observations $\mathcal{O}^t$ and proximity matrices $[\rho_{ij}^{\omega_t}]$ built on theses datasets. Assuming that the assumption of structural stability holds it follows that a minimal generating set $\mathcal{S}^{t+1} \subset \mathcal{O}^{t+1}$ exists which generates the same RF results as RF($\mathcal{O}^{t+1}$) using proximity based contagion and* *feedback with the proximity information at time t:*

$$RF([\sum_{k:\mathcal{O}_{k.}\in\mathcal{S}^{t+1}}[\rho_{ij}^{\omega_t}]_{ik}^{n}\mathcal{S}_{kj}^{t+1}]_{ij}) = RF([\sum_{k:\mathcal{O}_{k.}^{t+1}\in\mathcal{O}^{t+1}}[\rho_{ij}^{\omega_{t+1}}]_{ik}^{n}\mathcal{O}_{kj}^{t+1}]_{ij}) = RF(\mathcal{O}^{t+1}) \quad (11)$$

Proof. This follows directly from corollary 1 and the assumption of structural stability. □

Corollary 2 is the main result on this paper: under the outlined assumptions it is sufficient to estimate the future development of a small number of observations in order to estimate the future state of a dependent variable and the whole dataset. The result can be used for prediction in general, early warning or stress testing and allows by the concept of importance measurement to identify the main risk drivers of future event occurrence.

4 Empirical Study

The Dataset

To make full use of the capabilities of Random Forests a large number of independent variables/risk indicators should be used. Also the aim is to show that the proximity based stress testing framework can predict/warn about future crises, thus the used dataset should be a time series. Therefore the public and online available data of the World Bank, "World Development Indicators & Global Development Finance" have been sourced. [1].
The independent indicators are selected as in chapter **??** from currently applied theories on GDP growth, such as tax raising, public spending, monetary policy, the liberty of the economic environment, the workforce and its education and international trade. Indicators with more than 33% missing values are excluded. Indicators that cannot be easily compared between countries such as indicators measured in local currency or other absolute values are also not included. In numbers, 104 indicators are chosen between 1990 and 2010[2]. The large number of indicators in the model can easily be coped with by Random Forests and as (Biau 2012) shows, there will be no distortion from variables with no predictive power. The indicators and their descriptions are listed in the appendix **??** to the document. Since the Random Forests based recursive conditional partitioning does not over-fit (Breiman and Cutler 2005), many more indicators could theoretically be introduced.
The 20 years of data in the sample encompasses information from Australia, Austria, Belgium, Brazil, Canada, China, CzechRepublic, Denmark, Finland, France, Germany, Greece, HongKongSAR, Hungary, India, Indonesia, Ireland, Italy, Japan and the United States. The choice of countries to be included in the sample is very important (Pasricha et al. 2013). The countries represented in this study are mostly from the developed world but also encompass emerging economies. The specific choice of the countries is

[1] Online in internet: http://data.worldbank.org/indicator
[2] With an average of 9% missing values between 1999-2010.

based upon data availability and quality.

As dependent variable an indicator for financial stability was chosen: This paper is considering the changes in the number of non-performing loans per country as such. The non performing loans (NPL's) are studied in various scientific papers. Espinoza and Prasad (2010) describe NPL as key macroeconomic indicator for financial stability and investigate its feedback effects over a 3 year period. They especially find that financial institutions with a high NPL are very sensitive to macroeconomic stress. Likewise Vatansever and Hepsen (2013) argue that NPL is an important economic performance measure and apply a regression and cointegration analysis to show a significant relationship between NPLs and a list of macroeconomic indicators. Inaba et al. (2005) finally analyse the interrelationship between the increase in non-performing loans (NPLs) and the performance of the real economy in Japan modelling first the effect of macroeconomic variables on NPLs and then the respective feedback effect of a raise in NPLs on the economy finding significant distorting influence.

The number of non-performing loans as share of total loans as share of GDP is again a public and online available indicator of the World Bank, "World Development Indicators & Global Development Finance" [3].

Since the recursive conditional partitioning framework is used as a classification algorithm, the dependent variable has to be binary. An event, which is a rise in non-performing loans (NPL), is defined as whenever the respective NPL is higher than in the previous year. The dependent variable Y will take the value 1 for an adverse movements in NPL and 0 for other instances.

Commonly an event based on NPL movement occurs only after a certain threshold. For example when the ratio exceeds 20% (Pasricha et al. 2013). However, in this paper a threshold of 0 is chosen thus a rise in NPL is considered an event while this paper considers events in most of the countries in a single year as a stress scenario. Within the analysed years from 1999 to 2010 (in this paper a 10 year rolling window is applied, thus the analysis starts only in 1999) the NPLs in the sample evolve as shown in figure 1:

The NPL evolution shows a stress period in 2001/2002 and 2006 to 2009. This coincides with empirically observed crises. The proposed model will be applied to all scenarios in the sample, stress as well as calmer times. This paper claims that the model will be able to amply reproduce the share of rising NPL of each each period.

Parameters of the Random Forest-cforest Model

The cforest algorithm implemented in the R 'party' package is applied, using the following parameter settings: quadratic test-statistics with a splitting criterion of a variable which is associated with at least 99% significance, a minimum sum of weights in the knot of twice the weight of non-event cases and a minimum sum of weights of each of

[3] Online in internet: http://data.worldbank.org/indicator

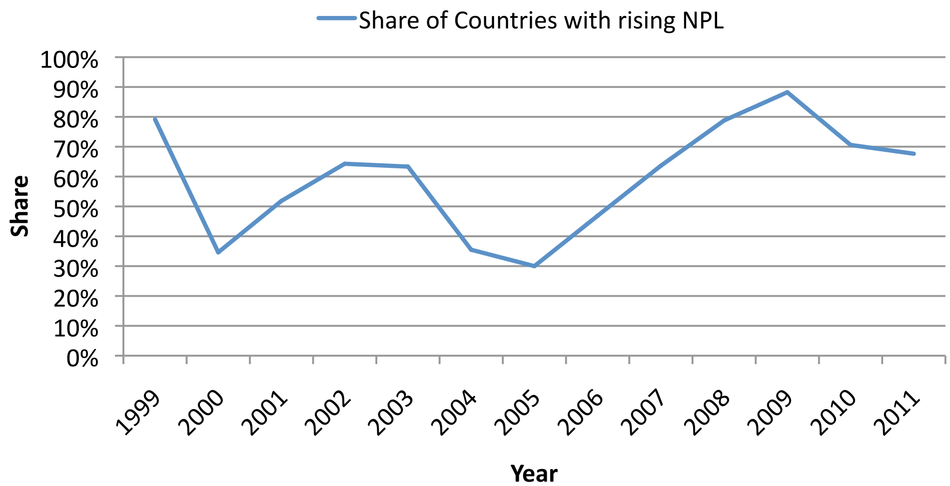

Figure 1: Evolution of share countries with rising NPL per year in the sample.

the subsequent knots of the weight of event cases. This choice is expected to lead to high accuracy but no empty knots are possible. The weights themselves are the inverse proportion of the number of events or non-events in the dataset. The number of sampled variables tried at each split is set to the square root of the number of independent variables. The choices with regard to sampled variables and weights are as suggested by(Breiman and Cutler 2005). For the stability of the results, 2,500-125,000 trees are run for each forest.

The analysis is done on a historic rolling window of 10 years. As implied by the theory in section 3 there will be no sampling and 100% of the data is used to built the trees. However (Strobl et al. 2007) suggest the bootstrap sample to be set to 63.2%. Since this application reduces overfitting it will be applied as well.

Process

To assess the performance of the proposed model, the following analysis tries to predict, for a specific point in time, the succeeding two years by starting with an RF model and proximity matrix on the preceding two years[4]. As laid out before, the framework does not specify a stress testing scenario. To test the proposed model and framework this paper will apply the described time series form the Worldbank database in the following way: The aim is to test whether the model can predict the whole future (stress) situation based on a stressed sample today. In an empirical dataset, seen in retrospective, a well chosen stress scenario at time t reflects the figures which will actually be incurred at time t+1. Thus if it should be shown that the proposed macro stress testing framework

[4]A time window of two years is the minimum in this paper in order to have sufficient data points to assure the quality of the results.

is able to correctly predict a future (stress) situation from a sample of stressed observations today, a sample of observations from time t+1 can be inserted in the model at time t and the prediction by proximity based contagion and feedback can be compared to the actually incurred number of events. For example: To predict the state of the economy in 2007/2008 by proximity based contagion and feedback, a RF model is drawn on the years 2005/2006. In the next step a sample of observations from the 2007/2008 dataset is chosen and the respective observations in the 'normal state' the 2005/2006 dataset are stressed by replacement by the chosen sample. Then the remaining observations in 2005/2006 are 'infected' by application of proximity based contagion: the values of the observations in each independent variable are replaced by the proximity weighted average of the respective values of the variables of the inserted observations. Thus on each variable or risk indicator j, the value of observation $\mathcal{O}_i, \mathcal{O}_{ij}$ is replaced by the proximity weighted average of the values on variable j of the observations in the stress sample. These averages differ from each other for each observation i since each un-stressed observation is likely to have different proximities to the observations in the stress sample. Note that the proximity based contagion is applied on the full observation thus likewise on the dependent variable Y^t. On the resulting estimated stress dataset a new Random Forests is drawn, predicting the stressed state of the economy. The samples observations are not replaced by averages since they are assumed to correctly reflect the future state.

In case feedback is included in the model, the proximity weighted averages to update each value of each variable are calculated on all observations, the stressed ones and the un-stressed ones. This because the basic idea of feedback is to model the effect of the changed observations on the initially inserted observations from the stressed sample. Feedback effects are often modelled for several years. See for example Espinoza and Prasad (2010). Thus the process is iterated in accordance with the number of years the feedback should be simulated for: If two years of feedback are simulated the proximity weighted averages are built twice, first on the initial values and second on the proximity weighted averages of the initial values. The iteration of the proximity based feedback and contagion is reflecting the intuition that the feedback of the effects of initially inserted stress sample observations is affecting all participants interdependently and that it is fading with time. The fading effect is a logical consequence of taking averages of averages.

Applied on all years in the empirical dataset this process will obviously assess the predictive power of the methodology in stressed as well as normal situations. Nonetheless the focus is on the crises, thus times of stress. As such, the training dataset will be called initial dataset, the resulting dataset from the application of the methodology of proximity based contagion and feedback will be called estimated stressed dataset and the drawn sample to initiate the methodology is the stress sample. The future dataset which is predicted and from which the stress sample is drawn is simply the future dataset.

Model Accuracy and Calibration

As mentioned, to assess the accuracy of the applied model a new forest is drawn on the resulting stressed dataset, including the before mentioned proximity weighted update of the dependent variable Y^t. Then $\hat{Y^t}$ is predicted using this new forest and the stressed data. The result is compared to the observed classification of Y^{t+1}. The accuracy is measured by three types of error: the type 1 error, the share of events which have been classified as non-events, the type two error, the share of non-events which have been classified as events and the average classification error of the two, the average error. In most of the Random Forests applications some form of the average error is reported.

In this study the the Random Forests-cf model is calibrated to minimize the average classification error of the fitted forests by specific calibration of the class weights. Class weights are again the inverse proportion of the number of events or non-events in the dataset (Breiman and Cutler 2005). Additionally it is preferable if the type 1 and type 2 errors are about the same size and consequently of the same size as the average classification error.

Samples

Based on the design of the proposed model it is most suitable to choose as stress sample from the future dataset those observations which are the most connected in the dataset, thus with the highest proximity measures: the observations are sorted with regard to the mode of their proximities and a predefined share is chosen from amongst the top entries of that list. The mode is taken as measure because an observation which is most often most highly connected to other observation is more contagious than an observation which has the highest mean, which could stem from a few close observations only. The empirical dataset revers to the total empirical data used in the analysis.

In accordance with the assumptions in proposition 1 a full training sample is assumed when building each tree, thus all observations, the initial and stressed ones are used.

4.1 Empirical Results

Before starting the analysis, proposition 1 is verified on the empirical dataset: on a subset of the empirical dataset a random forest is drawn. Afterwards a sample of observations is chosen and the remaining observations are replaced by the proximity weighted average in accordance with the method outlined above. On the resulting dataset a new random forest is drawn and the prediction of the events is compared to the prediction of the events of the initial random forest. Note that to verify proposition 1 the "stress" sample is not drown from the future dataset but from the same initial dataset that the random forest is drawn on. With this approach it is possible to show that with only a fraction of the initial dataset and proximity based contagion the same results can be achieved as with the whole initial dataset. The size of the sample varies from 0 to 100% to derive a flavour of how large a sample should be to derive an accurate approximation to the initial dataset.The following graph shows the evolution of the type 1 and type 2 errors in relation to the drawn sample size:

The used years are 2005-07, however the results are equivalent on other subsets. Using no sample and all data, class 1 has a classification error of 0 and class 0 a classification error of 6%: In other words the type 2 error is 0 and the type 1 error is 6%.

Figure 2: Type 1 and 2 Errors in Relation to the Sample initiating the Proximity Based Contagion Framework using the initial Dataset only

As stated in proposition 1 there is a minimal set producing the same Random Forest prediction as the whole dataset which in this case is 60% of the data (in this example the stress sample is randomly drawn and not based on an analysis of the mode). Both error types however remain relatively stable until the sample is reduced to 10% and below where the errors quickly rise to around 50% which basically means that the model assigns the classes randomly.

Based on this findings, the next step is a decision on how large the stress sample should be for the remainder of the analysis. Therefore the above analysis is repeated on a 2 year rolling window and with a comparison of stress sample size s of 10%, 33% and 50%. Note that again the stress sample is drawn on the respective initial dataset and not the future dataset since this analysis aims at finding a suitable sample size on known data and then test the model, including the chosen sample size, on unknown, out-of-the-sample data. The effects on the average error of the estimated forest are depicted in table 1.

Table 1: Results of Random Forests Prediction on a dataset built by proximity based contagion with full Tree Growing Sample and 10%, 33% or 50% Stress Sample Size

Years	Average-Error with 10% Sample Size	Average-Error with 33% Sample Size	Average-Error with 50% Sample Size
1999-2000	0.44	0.35	0.27
2000-2001	0.49	0.37	0.3
2001-2002	0.54	0.41	0.32
2002-2003	0.47	0.35	0.26
2003-2004	0.37	0.26	0.17
2004-2005	0.43	0.28	0.2
2005-2006	0.42	0.3	0.21
2006-2007	0.44	0.33	0.22
2007-2008	0.39	0.27	0.19
2008-2009	0.38	0.22	0.15
2009-2010	0.36	0.25	0.21
Average	0.43	0.31	0.23

The results show a persistent pattern though all years of reduced average-errors whenever more data is inserted. In their similar analysis Alessi and Detken (2014) derive a type 1 error of 38%, a type 2 error of 25% and thus an average-error 32%. Accordingly in this paper an average-error which is roughly below 33% is considered suitable. Based on the results in table 1 a stress sample of the size 33% of the dataset leads on average to an average error of 31%. Thus a stress sample size of 33% will be deployed throughout the paper.

4.1.1 Empirical Results

To assess the performance of the framework the average error, the type 1 and 2 error are calculated. Additionally it is assessed whether the application of proximity based contagion and feedback to the dataset adds accuracy or whether the application to the dependent variable Y only is sufficient. Thus the 3 errors are likewise calculated on a forest model on the estimated stressed Y with the initial unchanged dataset, as well as on a forest model on the whole estimated stressed dataset.
Following the detailed results of a proximity based stress testing with a full training sample and a 33% share of stressed initial observations. The analysis is done for two year windows, the table column 'years' shows the oldest year of the training dataset and the youngest of the predicted set, the average error of the estimated dataset is the average error of a Random Forests prediction on the estimated stressed dataset while the average error of the initial dataset is the average error of a Random Forests prediction on the initial, unchanged dataset:

Table 2: Results of Stress Forecasting by Proximity Based Contagion with Full Tree Growing Sample, 33% Stressed Inputs

Years	Average-Error Estimated Dataset	Average-Error of Initial Dataset	Estimated Data - Error Type 2	Estimated Data - Error Type 1	Initial Data - Error Type 2	Initial Data - Error Type 1
1999-2000	0.35	0.51	0.69	0	0.58	0.44
2000-2001	0.37	0.52	0.41	0.32	0.86	0.18
2001-2002	0.41	0.55	0.46	0.35	0.78	0.31
2002-2003	0.35	0.49	0.4	0.3	0.73	0.24
2003-2004	0.26	0.47	0.32	0.19	0.23	0.7
2004-2005	0.28	0.49	0.4	0.16	0.32	0.67
2005-2006	0.3	0.51	0.29	0.31	0.53	0.49
2006-2007	0.33	0.51	0.37	0.29	0.84	0.18
2007-2008	0.27	0.5	0.26	0.28	1	0.01
2008-2009	0.22	0.5	0.21	0.23	0.99	0.01
2009-2010	0.25	0.44	0.44	0.06	0.83	0.05

It can quickly be seen that the prediction using proximity based contagion results in an average-error around 30% which significantly improves the average error in comparison to the initial average error of 50%. Additionally the framework balances the initial type 1 and 2 errors, for example during the crisis period 2006-07, 2007-08 and 2008-09. It needs to be noted that the model is stable and accurate ahead of the crisis and during the crisis.

However, this model is based on contagion only. The stressed inputs transfer the shock by proximity weighted averages to the rest of the dataset. To introduce feedback effects, the process has to be iterated and all observations have to be included in the weighted average to cover feedback from already infected observations:

Table 3: Results of Stress Forecasting by Proximity Based Contagion with Full Tree Growing Sample, 33% Stressed Inputs and Feedback

Years	Average-Error Estimated Dataset	Average-Error of Initial Dataset	Estimated Data - Error Type 2	Estimated Data - Error Type 1	Initial Data - Error Type 2	Initial Data - Error Type 1
1999-2000	0.27	0.51	0.55	0	0.56	0.46
2000-2001	0.37	0.51	0.42	0.33	0.85	0.16
2001-2002	0.41	0.54	0.45	0.37	0.76	0.32
2002-2003	0.34	0.48	0.39	0.29	0.76	0.19
2003-2004	0.26	0.47	0.33	0.19	0.26	0.69
2004-2005	0.27	0.52	0.4	0.15	0.41	0.63
2005-2006	0.31	0.5	0.29	0.33	0.49	0.52
2006-2007	0.32	0.5	0.32	0.31	0.8	0.2
2007-2008	0.28	0.51	0.28	0.27	1	0.01
2008-2009	0.24	0.49	0.24	0.24	0.97	0.01
2009-2010	0.25	0.45	0.44	0.05	0.86	0.04

The results are very similar: the average of the average-errors of the estimated

dataset is in both cases roughly 30% (31% without feedback, 30% with feedback). On the other hand the volatility of the difference between the type 2 and type 1 error estimates decreases from 20% to 17%.
Although the theory requires the training sample for the trees to be full, it is worthwhile to check a smaller sample when dealing with real world data. After all, a full training sample enlarges the risk of overfitting. The next analysis employs thus the suggested training sample size of 63.2% (Hothorn et al. 2006a):

Table 4: Results of Stress Forecasting by Proximity Based Contagion with 63% Tree Growing Sample, 33% Stressed Inputs and Feedback

Years	Average-Error Estimated Forest	Average-Error of Initial Forest	Estimated Data - Error Type 2	Estimated Data - Error Type 1	Initial Data - Error Type 2	Initial Data - Error Type 1
1999-2000	0.26	0.51	0.41	0.1	0.38	0.63
2000-2001	0.35	0.51	0.29	0.42	0.95	0.06
2001-2002	0.41	0.6	0.47	0.34	0.82	0.39
2002-2003	0.39	0.5	0.39	0.4	0.96	0.04
2003-2004	0.31	0.45	0.37	0.25	0.67	0.22
2004-2005	0.41	0.5	0.4	0.41	0.55	0.44
2005-2006	0.34	0.48	0.3	0.38	0.27	0.68
2006-2007	0.33	0.49	0.35	0.32	0.67	0.31
2007-2008	0.39	0.5	0.39	0.39	1	0
2008-2009	0.34	0.46	0.4	0.27	0.93	0
2009-2010	0.28	0.48	0.45	0.1	0.93	0.04

The average-error on the estimated dataset has increased to 34% while the volatility of the difference between type 2 and type 1 errors decreased significantly to 11%. Thus basing the method on a smaller training sample for each tree leads to slightly less accuracy but lower differences between the type 1 and 2 classification error estimates. Depending on whether additional data is supposed to be inserted in a once built forest either the full training sample or the version with 63% tree growing sample size might be suitable.

4.2 Policy Indicators

The application of the Random Forests model allows to assess the most important variables within the proximity based stress testing framework. This adds value in the following way: The proximity based contagion and feedback model allows to identify beforehand periods of stress. Once identified the most important variables, key risk drivers, with regard to the stress period can be established. Thus the model shows, contingent on the correctness of the Random Forests, which risk drivers are important in a coming stress event and thus which variables could be changed to prevent the results from the scenario. In other words the model points out policy indicators.

To test now whether variables which are identified as important in the estimated dataset are also important in the real future stressed dataset, the upper percentile of the distribution of the importance score of both datasets are compared and it is assessed how many variables in the same percentile are actually the same. This shows whether the application of the above outline methodology leads to the same importance ranking of variables in the estimated stress situation as in the actually observed stress situation.

To calculate the importance score, the importance measure introduced within the recursive conditional partitioning package (party, cforest) is applied (Hothorn et al. 2006a). It is defined in the following way: *Importance is defined by randomly permuting the values of a predictor variable and thus breaking its original association with the response. Thus, a reasonable measure for variable importance is the difference in prediction accuracy before and after permuting a variable, averaged over all trees* (Strobl et al. 2009).

Table 5: Persistence of Most Important Variables in Stress Estimation and Empirical Data, by Percentile of the Importance Distribution

Percentile	10%	15%	20%	25%	33%
Share of Variables which are Important in both Datasets' Percentiles	0.3	0.53	0.63	0.67	0.63

In a proximity based stress testing framework with a training sample size of 63.2%, a stress sample of the size of 33% and feedback effects, around 33%-67% of the most important variables in the estimated dataset also contribute to the future stressed data. This is sufficient to state that important variables from the stress testing exercise actually are important in a crisis situation.
Following in more detail, which variables are actually important in both datasets:

Table 6: List of Most Important Variables and their Persistence in the Stress Scenario

Percentile	Variable Short Name	Variable Description	Occurence
	FM.LBL.MQMY.IR.ZS	Money and quasi money (M2) to total reserves ratio	1
	TX.VAL.MRCH.RS.ZS	Merchandise exports by the reporting economy, residual (% total merchandise exports)	1
	PA.NUS.ATLS	DEC alternative conversion factor (LCU per USUSD)	0
	MS.MIL.XPND.ZS	Military expenditure (% of central government expenditure)	1
	GC.TAX.TOTL.GD.ZS	Tax revenue (% of GDP)	1
	IP.PAT.RESD	Patent applications, residents	0
	GC.XPN.TOTL.GD.ZS	Expense (% of GDP)	1
	TX.VAL.FOOD.ZS.UN	Food exports (% of merchandise exports)	1
	NY.GDP.DEFL.ZS	GDP deflator (base year varies by country)	0
0.1	FP.CPI.TOTL	Consumer price index (2005 = 100)	1
	TX.VAL.MMTL.ZS.UN	Ores and metals exports (% of merchandise exports)	0
	NY.GNS.ICTR.ZS	Gross savings (% of GDP)	1
	NE.CON.GOVT.ZS	General government final consumption expenditure (% of GDP)	1
	IP.TMK.NRES	Trademark applications, direct nonresident	1
0.15	CM.MKT.LCAP.GD.ZS	Market capitalization of listed companies (% of GDP)	1
	NY.GDP.MKTP.KD.ZG	GDP growth (annual %)	1
	NE.EXP.GNFS.KD.ZG	Exports of goods and services (annual % growth)	1
	SH.XPD.PUBL	Health expenditure, public (% of total health expenditure)	0
0.2	IS.AIR.GOOD.MT.K1	Air transport, freight (million ton-km)	0
	TM.VAL.MRCH.XD.WD	Import value index (2000 = 100)	1
	TX.QTY.MRCH.XD.WD	Export volume index (2000 = 100)	0
	FM.LBL.MQMY.ZG	Money and quasi money growth (annual %)	1
	IP.PAT.NRES	Patent applications, nonresidents	0
0.25	NE.GDI.TOTL.KD.ZG	Gross capital formation (annual % growth)	1
	FD.RES.LIQU.AS.ZS	Bank liquid reserves to bank assets ratio (%)	1
	TM.TAX.MRCH.SM.AR.ZS	Tariff rate, applied, simple mean, all products (%)	0
	NE.IMP.GNFS.KD.ZG	Imports of goods and services (annual % growth)	1
	IS.AIR.DPRT	Air transport, registered carrier departures worldwide	0
	GC.TAX.GSRV.VA.ZS	Taxes on goods and services (% value added of industry and services)	1
	SE.XPD.TOTL.GD.ZS	Public spending on education, total (% of GDP)	0
	ST.INT.DPRT	International tourism, number of departures	0
0.33	TM.VAL.FOOD.ZS.UN	Food imports (% of merchandise imports)	1

The most important shared indicators are money and quasi money (M2) to total reserves ratio, Merchandise exports by the reporting economy, Military expenditure, Tax Revenue and Expense.

Indicators like money and quasi money (M2) to total reserves ratio or tax revenue and expense can be called direct policy indicators. The reserve ratio can be changed by central banks and has an effect for example on money supply and the interest rate. Likewise if taxation is identified as an influential indicator it can be used directly as a policy indicator. On the other hand, many of the indicators, alike military expenditures, are not direct policy indicators. Although other scholars have observed that military spending might go up in certain countries in times of a crisis, lowering it would probably not help. Thus either the included variables are direct policy indicators like tax revenue and expenses or an indirect theory behind a risk indicator can be used.

The implementation of the proximity based stress testing framework as a whole can thus be summarised as follows:

1. On todays data a Random Forests is drawn and a proximity matrix is built.
2. The most interrelated observations are stressed

3. The future dataset is estimated by proximity based contagion and feedback
4. The future events are estimated.
5. If there is no significant increase in events, the dataset is not affected by the chosen stress scenario. If on the other hand there is a significant increase in events, the proximity based framework can be used to identify policy indicators to address the weaknesses in the dataset before the crisis assumed on the stress scenario emerges
6. Identify the most important variables on the estimated future dataset.
7. Translate the most important variables into policy indications

5 Conclusion

In this a paper a non-linear macro-stress testing methodology, the proximity based stressed testing framework, with focus on early warning and crisis remediation was developed. The development was done based on heuristic derivation and mathematical proofs. The proposed methodology builds on a conditional recursive partitioning forest: by application of its proximity measures, the effects of a small stressed sample are expanded to the whole dataset. If feedback effects are required the process is iterated. It was shown that a Random Forest model on the estimated stressed dataset predicts a potential crisis very well by accurately forecasting the number of events. Likewise it has been shown that the most important variables leading to this events can be identified and used as input to manage or prevent crises.

In comparison to the no model and the similar model of Alessi and Detken (2014) the proposed achieved a lower average and type 1 errors:

Table 7: Comparison of the Results of the Proposed Proximity based Stress Testing Framework against no Model use and a suitable Benchmark

	Time Window	Average-Error	Error Type 1	Error Type 2
Proximity based stress testing Framework	1999-2010	0.3	0.23	0.37
(33% stress sample, full training sample, feedback)	2005-2009	0.29	0.29	0.28
No Model	1999-2010	0.5	0.29	0.7
	2005-2009	0.5	0.19	0.82
Benchmark Alessi and Detken (2014)	1970-2013	0.32	0.38	0.25

Especially during the years of the crisis the proximity based stress testing framework exhibits a low average classification error and almost equal type 1 and 2 errors.

The proposed proximity based stress testing framework is designed to consider most requirements formulated by the BIS (Borio et al. 2012) such as being non-linear, naturally defined contagion and feedback effects and the capability to incorporate national and international KRIs. However, initially the framework still relies on historical data. With regard to the BIS critics towards the application of early warning systems, the proposed framework addresses this by not identifying early warning indicators but by modelling the stressed events using the interaction between all included indicators and then identifying indicators for crises remediation. The number of the modelled stress events is the early warning indicator.

Due to the inherited characteristics of Random Forests it is compatible with the application of big data.

References

Aikman, D., et al., 2009: Funding liquidity risk in a quantitative model of systemic stability. *Bank of England Working Paper*, **372**.

Alessi, L. and C. Detken, 2014: Identifying excessive credit growth and leverage. *ECB Working Paper Series*, **1723**.

Allen, F. and D. Gale, 2000: Financial contagion. *The Journal of Political Economy*, **108**, 1–33.

Babecky, J., T. Havranek, J. Mateju, M. Rusnak, K. Smidkova, and B. Vasicek, 2012: Leading indicators of crisis incidence, evidence from developed countries. *ECB Working Paper Series*, **1486**.

Barnhill, T. and L. Schumacher, 2011: Modeling correlated systemic liquidity and solvency risk in a financial environment with incomplete information. *IMF Working Paper*, **WP/11/263.**

Biau, G., 2012: Analysis of a random forests model. *Journal of Machine Learning Research*, **13**, 1063–1095.

Borio, C. and M. Drehmann, 2009: Assessing the risk of banking crises revisited. *BIS Quarterly Review*, **March**, 2946.

Borio, C., M. Drehmann, and K. Tsatsaronis, 2012: Bis working papers no 369 stress-testing macro stress testing: does it live up to expectations? *Online in internet: http://www.bis.org/publ/work369.htm (Access: 30.06.15).*

Boss, M., et al., 2008: Stress tests for the austrian fsap update 2007: Methodology, scenarios and results. *OENB Financial Stability Report*, **15**.

Breiman, L. and A. Cutler, 2005: Random forests. *Available online at: http://www.stat.berkeley.edu/users/breiman/ RandomForests/cc_home.htm (accessed 25.07.2005).*

Breuer, T. and G. Krenn, 1999: Leitfaden zum marktrisiko band 5 durchfhrung von krisentests. *Online in internet: http://www.oenb.at/en/img/band5dv40_tcm16-11166.pdf*, **5**.

De Bandt, O. and P. Hartmann, 2001: *Financial crisis, contagion and the lender of last resort: A book of readings*, chap. Systemic risk: A survey, pp 249–98. Goodhart, C., Illing, G., Oxford University Press, London.

Drehmann, M., A. Patton, and S. Sorensen, 2007: Non-linearities and stress testing. *in Risk measurement and systemic risk, Proceedings of the fourth joint central bank research conference, ECB.*

Drehmann, M., S. Sorensen, and M. Stringa, 2010: The integrated impact of credit and interest rate risk on banks: A dynamic framework and stress testing application. *Journal of Banking & Finance*, **34**, pp 73551.

Drehmann, M. and N. Tarashev, 2011: Systemic importance: Some simple indicators. *BIS Quarterly Review*, **March**, 2537.

Duttagupta, R. and P. Cashin, 2011: Anatomy of banking crises in developing and emerging market countries. *Journal of International Money and Finance*, **30(2)**, 354–376.

Elsinger, H., A. Lehar, and M. Summer, 2006: Risk assessment for banking systems. *Management Science*, **52(9)**, 1301–41.

Espinoza, R. and A. Prasad, 2010: Nonperforming loans in the gcc banking system and their macroeconomic effects. *IMF Working Paper, Available online at: https://www.imf.org/external/pubs/ft/wp/2010/wp10224.pdf (accessed 17.09.2015).*

Frankel, J. A. and S. Wei, 2004: Managing macroeconomic crises. *NBER Working Papers*, **10907**.

Ghosh, S. R. and A. Ghosh, 2002: Structural vulnerabilities and currency crises. *IMF Staff Papers*, **50(3)**.

Hothorn, T., K. Hornik, and A. Zeileis, 2006a: party: A laboratory for recursive partytioning. *Available online at: http://cran.r-project.org/web/packages/party/vignettes/party.pdf (08.01.2015).*

Hothorn, T., K. Hornik, and A. Zeileis, 2006b: Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, **15, 3**, 651–674.

Inaba, N., T. Kozu, T. Sekine, and T. Nagahata, 2005: Non-performing loans and the real economy: Japans experience. *BIS Papers, Available online at: http://www.bis.org/publ/bppdf/bispap22g.pdf (accessed 27.09.2015)*, **22**.

Jacobsen, T., L. J., and K. Roszbach, 2005: Exploring interactions between real activity and the financial stance. *Journal of Financial Stability*, **1, 3**, 30841.

Juselius, M. and M. Kim, 2011: Sustainable financial obligations and crisis cycles. *Helsinki Economic Centre of Research Discussion Papers*, **313**.

Manasse, P. and N. Roubini, 2009: Rules of thumb for sovereign debt crises. *Journal of International Economics,*, **78**, 192–205.

Manasse, P., R. Savona, and M. Vezzoli, 2013: Rules of thumb for banking crises in emerging markets. *Universit di Bologna DSE Working Paper*, **872**.

Pasricha, G., T. Roberts, C. I., and B. Howell, 2013: Assessing financial system vulnerabilities: An early warning approach. *Online in internet: http://www.bankofcanada.ca/wp-content/uploads/2013/11/boc-review-autumn13-pasricha.pdf.*

Qu, Z., J. Wang, and R. A. Hull, 2005: Products of row stochastic matrices and their applications to cooperative control for autonomous mobile robots. *Proceedings of American Control Conference, Portland, Oregon*, **June**.

Savona, R. and M. Vezzoli, 2015: Fitting and forecasting sovereign defaults using multiple risk signals. *Oxford Bulletin of Economics and Statistics, forthcoming.*

Strasser, H. and C. Weber, 1999: On the asymptotic theory of permutation statistics. *Mathematical Methods of Statistics. Available online at: http://epub.wu-wien.ac.at/dyn/openURL?id=oai:epub. wu-wien.ac.at:epub-wu-01_94c.*, **8**, 220–250.

Strobl, C., A. Boulesteix, A. Zeileis, and T. Hothorn, 2007: Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, **8**, 1471–2105.

Strobl, C., J. Malley, and G. Tutz, 2009: An introduction to recursive partitioning: Rationale, application and characteristics of classification and regression trees, bagging and random forests. *Psychol Methods*, **14(4)**, 323348.

Upper, C. and A. Worms, 2004: Estimating bilateral exposures in the german interbank market: Is there a danger of contagion. *European Economic Review*, **48**, pp 827–49.

Van den End, W., 2008: Liquidity stress-tester: A macro model for stress-testing banks liquidity risk. *DNB Working Paper*, **175**.

Vatansever, M. and A. Hepsen, 2013: Determining impacts on non-performing loan ratio in turkey. *Journal of Finance and Investment Analysis*, **2(4)**, 119–129.

Waelchli, B., 2015: A random forests based structural performance ratio for credit portfolio management and optimisation. *Available at SSRN: http://ssrn.com/abstract=2550072 or http://dx.doi.org/10.2139/ssrn.2550072.*

A List of applied independent risk indicators

Table 8: List of Applied Risk Indicators 1/4

Variable Name	Variable Description	Theory Class
IC.REG.COST.PC.ZS	Cost of business start-up procedures (% of GNI per capita)	Economic Environment
PA.NUS.ATLS	DEC alternative conversion factor (LCU per US$)	Monetary
NE.IMP.GNFS.KD.ZG	Imports of goods and services (annual % growth)	International Trade
NE.GDI.TOTL.KD.ZG	Gross capital formation (annual % growth)	Spending
NE.GDI.TOTL.ZS	Gross capital formation (% of GDP)	Spending
NE.EXP.GNFS.KD.ZG	Exports of goods and services (annual % growth)	International Trade
FP.CPI.TOTL	Consumer price index (2005 = 100)	Macroeconomic
FP.CPI.TOTL.ZG	Inflation, consumer prices (annual %)	Macroeconomic
NE.CON.TETC.KD.ZG	Final consumption expenditure, etc. (annual % growth)	Spending
NE.GDI.FTOT.ZS	Gross fixed capital formation (% of GDP)	Spending
IC.ISV.DURS	Time to resolve insolvency (years)	Economic Environment
NY.GDP.DEFL.ZS	GDP deflator (base year varies by country)	Macroeconomic
NY.GDP.DEFL.KD.ZG	Inflation, GDP deflator (annual %)	Macroeconomic
IC.LGL.PROC	Procedures to enforce a contract (number)	Economic Environment
TM.VAL.FOOD.ZS.UN	Food imports (% of merchandise imports)	International Trade
TM.VAL.MRCH.XD.WD	Import value index (2000 = 100)	International Trade
NE.CON.GOVT.ZS	General government final consumption expenditure (% of GDP)	Spending
NY.GNS.ICTR.ZS	Gross savings (% of GDP)	Economic Environment
NY.GNS.ICTR.GN.ZS	Gross savings (% of GNI)	Economic Environment
FM.LBL.MQMY.GD.ZS	Money and quasi money (M2) as % of GDP	Monetary
TX.VAL.MRCH.XD.WD	Export value index (2000 = 100)	International Trade
IT.NET.USER.P2	Internet users (per 100 people)	Economic Environment
SL.EMP.1524.SP.MA.ZS	Employment to population ratio, ages 15-24, male (%)	Workforce and Education
TM.QTY.MRCH.XD.WD	Import volume index (2000 = 100)	International Trade
SH.XPD.PCAP.PP.KD	Health expenditure per capita, PPP (constant 2005 international $)	Spending
GB.XPD.RSDV.GD.ZS	Research and development expenditure (% of GDP)	Spending
IP.JRN.ARTC.SC	Scientific and technical journal articles	Spending
FM.LBL.MQMY.ZG	Money and quasi money growth (annual %)	Monetary
GC.TAX.TOTL.GD.ZS	Tax revenue (% of GDP)	Tax
BX.TRF.PWKR.DT.GD.ZS	Workers' remittances and compensation of employees, received (% of GDP)	Economic Environment
GC.XPN.TOTL.GD.ZS	Expense (% of GDP)	Spending
SL.EMP.TOTL.SP.MA.ZS	Employment to population ratio, 15+, male (%)	Workforce and Education
NE.GDI.TOTL.CD	Gross capital formation (current US$)	Spending
NE.DAB.TOTL.ZS	Gross national expenditure (% of GDP)	Spending
IT.CEL.SETS.P2	Mobile cellular subscriptions (per 100 people)	Economic Environment
NE.RSB.GNFS.ZS	External balance on goods and services (% of GDP)	International Trade
NY.GNS.ICTR.CD	Gross savings (current US$)	Economic Environment
TX.VAL.TECH.CD	High-technology exports (current US$)	International Trade
NY.GDS.TOTL.CD	Gross domestic savings (current US$)	Economic Environment
BX.KLT.DINV.CD.WD	Foreign direct investment, net inflows (BoP, current US$)	International Trade

Table 9: List of Applied Risk Indicators 2/4

Variable Name	Variable Description	Theory Class
IC.REG.DURS	Time required to start a business (days)	Economic Environment
NY.GDS.TOTL.ZS	Gross domestic savings (% of GDP)	Spending
IC.LGL.DURS	Time required to enforce a contract (days)	Economic Environment
NE.GDI.FTOT.CD	Gross fixed capital formation (current US$)	Spending
BM.KLT.DINV.GD.ZS	Foreign direct investment, net outflows (% of GDP)	International Trade
IC.REG.PROC	Start-up procedures to register a business (number)	Economic Environment
IS.AIR.GOOD.MT.K1	Air transport, freight (million ton-km)	Economic Environment
TX.VAL.MRCH.WL.CD	Merchandise exports by the reporting economy (current US$)	International Trade
SH.XPD.PCAP	Health expenditure per capita (current US$)	Spending
GC.XPN.TRFT.ZS	Subsidies and other transfers (% of expense)	Spending
SL.EMP.1524.SP.ZS	Employment to population ratio, ages 15-24, total (%)	Workforce and Education
SH.XPD.PUBL.ZS	Health expenditure, public (% of GDP)	Spending
TM.VAL.MANF.ZS.UN	Manufactures imports (% of merchandise imports)	International Trade
NY.TAX.NIND.CD	Net taxes on products (current US$)	Tax
TX.VAL.MRCH.CD.WT	Merchandise exports (current US$)	International Trade
TM.VAL.MRCH.CD.WT	Merchandise imports (current US$)	International Trade
BN.CAB.XOKA.GD.ZS	Current account balance (% of GDP)	International Trade
GC.REV.SOCL.ZS	Social contributions (% of revenue)	Tax
GC.XPN.COMP.ZS	Compensation of employees (% of expense)	Economic Environment
NE.GDI.TOTL.KD	Gross capital formation (constant 2000 US$)	Spending
NE.IMP.GNFS.CD	Imports of goods and services (current US$)	International Trade
NE.GDI.FTOT.KD	Gross fixed capital formation (constant 2000 US$)	Spending
GC.XPN.INTP.ZS	Interest payments (% of expense)	Spending
BX.KLT.DINV.WD.GD.ZS	Foreign direct investment, net inflows (% of GDP)	International Trade
SH.XPD.PUBL.GX.ZS	Health expenditure, public (% of government expenditure)	Spending
BM.GSR.TRAN.ZS	Transport services (% of service imports, BoP)	International Trade
NE.EXP.GNFS.CD	Exports of goods and services (current US$)	International Trade
NE.CON.PETC.ZS	Household final consumption expenditure, etc. (% of GDP)	Economic Environment
SL.EMP.TOTL.SP.ZS	Employment to population ratio, 15+, total (%)	Workforce and Education
SL.EMP.1524.SP.FE.ZS	Employment to population ratio, ages 15-24, female (%)	Workforce and Education
TX.VAL.MRCH.R5.ZS	Merchandise exports to developing economies in South Asia (% of total merchandise exports)	International Trade
NE.RSB.GNFS.CD	External balance on goods and services (current US$)	International Trade
SE.PRM.ENRL.FE.ZS	Primary education, pupils (% female)	Workforce and Education
TM.VAL.TRAN.ZS.WT	Transport services (% of commercial service imports)	International Trade
GC.TAX.OTHR.RV.ZS	Other taxes (% of revenue)	Tax
SH.XPD.TOTL.ZS	Health expenditure, total (% of GDP)	Spending
BN.GSR.MRCH.CD	Net trade in goods (BoP, current US$)	International Trade
TX.QTY.MRCH.XD.WD	Export volume index (2000 = 100)	International Trade
TX.VAL.MANF.ZS.UN	Manufactures exports (% of merchandise exports)	International Trade
NE.CON.TETC.ZS	Final consumption expenditure, etc. (% of GDP)	Spending
SE.PRM.AGES	Primary school starting age (years)	Workforce and Education
IP.TMK.TOTL	Trademark applications, total	Economic Environment
TT.PRI.MRCH.XD.WD	Net barter terms of trade index (2000 = 100)	International Trade
NY.GSR.NFCY.CD	Net income from abroad (current US$)	International Trade
SL.TLF.CACT.MA.ZS	Labor participation rate, male (% of male population ages 15+)	Workforce and Education
NE.CON.PRVT.PC.KD	Household final consumption expenditure per capita (constant 2000 US$)	Economic Environment
FS.AST.DOMS.GD.ZS	Domestic credit provided by banking sector (% of GDP)	Economic Environment
ST.INT.TRNR.CD	International tourism, receipts for passenger transport items (current US$)	International Trade
TM.VAL.OTHR.ZS.WT	Computer, communications and other services (% of commercial service imports)	International Trade
EG.ELC.LOSS.KH	Electric power transmission and distribution losses (kWh)	Economic Environment
TX.VAL.MRCH.R3.ZS	Merchandise exports to developing economies in Latin America and the Caribbean (% of total merchandise exports)	International Trade
TM.VAL.MMTL.ZS.UN	Ores and metals imports (% of merchandise imports)	International Trade
SE.XPD.TOTL.GD.ZS	Public spending on education, total (% of GDP)	Spending
TX.VAL.MRCH.R6.ZS	Merchandise exports to developing economies in Sub-Saharan Africa (% of total merchandise exports)	International Trade

Table 10: List of Applied Risk Indicators 3/4

Variable Name	Variable Description	Theory Class
TX.VAL.MRCH.RS.ZS	Merchandise exports by the reporting economy, residual (% of total merchandise exports)	International Trade
ST.INT.TVLX.CD	International tourism, expenditures for travel items (current US$)	International Trade
SE.ENR.PRIM.FM.ZS	Ratio of female to male primary enrollment (%)	Workforce and Education
NE.EXP.GNFS.ZS	Exports of goods and services (% of GDP)	International Trade
BN.GSR.GNFS.CD	Net trade in goods and services (BoP, current US$)	International Trade
ST.INT.XPND.MP.ZS	International tourism, expenditures (% of total imports)	International Trade
TX.VAL.MRCH.HI.ZS	Merchandise exports to high-income economies (% of total merchandise exports)	International Trade
TX.VAL.FOOD.ZS.UN	Food exports (% of merchandise exports)	International Trade
SE.SEC.ENRL.GC.FE.ZS	Secondary education, general pupils (% female)	Workforce and Education
TX.VAL.TECH.MF.ZS	High-technology exports (% of manufactured exports)	International Trade
NE.CON.GOVT.KD.ZG	General government final consumption expenditure (annual % growth)	Spending
BN.KLT.DINV.CD	Foreign direct investment, net (BoP, current US$)	International Trade
SE.PRM.ENRR.FE	School enrollment, primary, female (% gross)	Workforce and Education
NE.IMP.GNFS.ZS	Imports of goods and services (% of GDP)	International Trade
SH.XPD.PRIV.ZS	Health expenditure, private (% of GDP)	Spending
BM.GSR.FCTY.CD	Income payments (BoP, current US$)	Economic Environment
FS.AST.DOMO.GD.ZS	Claims on other sectors of the domestic economy (% of GDP)	Economic Environment
BN.TRF.KOGT.CD	Net capital account (BoP, current US$)	International Trade
NY.TRF.NCTR.CD	Net current transfers from abroad (current US$)	International Trade
NE.CON.PRVT.CD	Household final consumption expenditure (current US$)	Economic Environment
BX.GSR.FCTY.CD	Income receipts (BoP, current US$)	Economic Environment
NE.CON.TETC.CD	Final consumption expenditure, etc. (current US$)	Spending
BX.GSR.TRVL.ZS	Travel services (% of service exports, BoP)	International Trade
BM.TRF.PWKR.CD.DT	Workers' remittances and compensation of employees, paid (current US$)	Economic Environment
NE.CON.PETC.CD	Household final consumption expenditure, etc. (current US$)	Economic Environment
SE.SEC.ENRL.FE.ZS	Secondary education, pupils (% female)	Workforce and Education
IT.MLT.MAIN.P2	Telephone lines (per 100 people)	Economic Environment
NE.CON.PRVT.PP.KD	Household final consumption expenditure, PPP (constant 2005 international $)	Economic Environment
NE.EXP.GNFS.KD	Exports of goods and services (constant 2000 US$)	International Trade
SL.TLF.CACT.FE.ZS	Labor participation rate, female (% of female population ages 15+)	Workforce and Education
TX.VAL.TRAN.ZS.WT	Transport services (% of commercial service exports)	International Trade
ST.INT.XPND.CD	International tourism, expenditures (current US$)	International Trade
NE.CON.TOTL.CD	Final consumption expenditure (current US$)	Spending
NE.DAB.TOTL.KD	Gross national expenditure (constant 2000 US$)	Spending
TM.VAL.INSF.ZS.WT	Insurance and financial services (% of commercial service imports)	International Trade
BN.CAB.XOKA.CD	Current account balance (BoP, current US$)	International Trade
TX.VAL.MRCH.OR.ZS	Merchandise exports to developing economies outside region (% of total merchandise exports)	International Trade
NE.IMP.GNFS.KD	Imports of goods and services (constant 2000 US$)	International Trade
FS.AST.CGOV.GD.ZS	Claims on central government, etc. (% GDP)	Economic Environment
GC.TAX.YPKG.ZS	Taxes on income, profits and capital gains (% of total taxes)	Tax
SE.PRM.ENRL	Primary education, pupils	Workforce and Education
IT.CEL.SETS	Mobile cellular subscriptions	Economic Environment
TX.VAL.TRVL.ZS.WT	Travel services (% of commercial service exports)	International Trade
IS.AIR.DPRT	Air transport, registered carrier departures worldwide	Economic Environment
NE.DAB.TOTL.CD	Gross national expenditure (current US$)	Spending
ST.INT.TRNX.CD	International tourism, expenditures for passenger transport items (current US$)	International Trade
BX.GSR.TOTL.CD	Exports of goods, services and income (BoP, current US$)	International Trade
SH.XPD.PUBL	Health expenditure, public (% of total health expenditure)	Spending
GC.XPN.OTHR.ZS	Other expense (% of expense)	Spending
NE.CON.GOVT.CD	General government final consumption expenditure (current US$)	Spending
BX.TRF.PWKR.CD.DT	Workers' remittances and compensation of employees, received (current US$)	Economic Environment
SE.SEC.AGES	Secondary school starting age (years)	Workforce and Education
SL.TLF.TOTL.FE.ZS	Labor force, female (% of total labor force)	Workforce and Education
BX.GSR.GNFS.CD	Exports of goods and services (BoP, current US$)	International Trade
NE.TRD.GNFS.ZS	Trade (% of GDP)	International Trade
TG.VAL.TOTL.GD.ZS	Merchandise trade (% of GDP)	International Trade

Table 11: List of Applied Risk Indicators 4/4

Variable Name	Variable Description	Theory Class
NE.CON.PETC.KD	Household final consumption expenditure, etc. (constant 2000 US$)	Spending
TX.VAL.MRCH.AL.ZS	Merchandise exports to economies in the Arab World (% of total merchandise exports)	International Trade
TX.VAL.MRCH.R4.ZS	Merchandise exports to developing economies in Middle East and North Africa (% of total merchandise exports)	International Trade
BX.TRF.CURR.CD	Current transfers, receipts (BoP, current US$)	International Trade
ST.INT.TVLR.CD	International tourism, receipts for travel items (current US$)	International Trade
TX.VAL.MMTL.ZS.UN	Ores and metals exports (% of merchandise exports)	International Trade
ST.INT.DPRT	International tourism, number of departures	International Trade
TM.VAL.SERV.CD.WT	Commercial service imports (current US$)	International Trade
BN.TRF.CURR.CD	Net current transfers (BoP, current US$)	International Trade
SE.SEC.ENRL.GC	Secondary education, general pupils	Workforce and Education
ST.INT.RCPT.CD	International tourism, receipts (current US$)	International Trade
TX.VAL.OTHR.ZS.WT	Computer, communications and other services (% of commercial service exports)	International Trade
GC.TAX.GSRV.RV.ZS	Taxes on goods and services (% of revenue)	Tax
BM.GSR.NFSV.CD	Service imports (BoP, current US$)	International Trade
NE.CON.PRVT.KD	Household final consumption expenditure (constant 2000 US$)	Spending
FS.AST.PRVT.GD.ZS	Domestic credit to private sector (% of GDP)	Economic Environment
BM.GSR.GNFS.CD	Imports of goods and services (BoP, current US$)	International Trade
IT.MLT.MAIN	Telephone lines	Economic Environment
BM.TRF.PRVT.CD	Private current transfers, payments (BoP, current US$)	International Trade
BX.PEF.TOTL.CD.WD	Portfolio equity, net inflows (BoP, current US$)	International Trade
SL.TLF.CACT.ZS	Labor participation rate, total (% of total population ages 15+)	Workforce and Education
BX.GSR.MRCH.CD	Goods exports (BoP, current US$)	International Trade
MS.MIL.XPND.GD.ZS	Military expenditure (% of GDP)	Spending
BM.GSR.TRVL.ZS	Travel services (% of service imports, BoP)	International Trade
GC.XPN.GSRV.ZS	Goods and services expense (% of expense)	Spending
SL.EMP.TOTL.SP.FE.ZS	Employment to population ratio, 15+, female (%)	Workforce and Education
BX.GSR.NFSV.CD	Service exports (BoP, current US$)	International Trade
ST.INT.RCPT.XP.ZS	International tourism, receipts (% of total exports)	International Trade
SE.PRM.ENRR	School enrollment, primary (% gross)	Workforce and Education
SL.TLF.TOTL.IN	Labor force, total	Workforce and Education
TX.VAL.SERV.CD.WT	Commercial service exports (current US$)	International Trade
NE.CON.TETC.KD	Final consumption expenditure, etc. (constant 2000 US$)	Spending
BX.GSR.TRAN.ZS	Transport services (% of service exports, BoP)	International Trade
BM.GSR.MRCH.CD	Goods imports (BoP, current US$)	International Trade
NE.CON.GOVT.KD	General government final consumption expenditure (constant 2000 US$)	Spending
BG.GSR.NFSV.GD.ZS	Trade in services (% of GDP)	International Trade
BM.GSR.TOTL.CD	Imports of goods, services and income (BoP, current US$)	International Trade
BX.GSR.CMCP.ZS	Communications, computer, etc. (% of service exports, BoP)	International Trade
TM.VAL.MRCH.AL.ZS	Merchandise imports from economies in the Arab World (% of total merchandise imports)	International Trade
ST.INT.ARVL	International tourism, number of arrivals	International Trade
SE.PRM.ENRR.MA	School enrollment, primary, male (% gross)	Workforce and Education
TM.VAL.TRVL.ZS.WT	Travel services (% of commercial service imports)	International Trade
NE.CON.PRVT.PP.CD	Household final consumption expenditure, PPP (current international $)	Economic Environment
SE.SEC.DURS	Secondary education, duration (years)	Workforce and Education
BN.GSR.FCTY.CD	Net income (BoP, current US$)	Economic Environment
SE.PRM.DURS	Primary education, duration (years)	Workforce and Education
BN.KAC.EOMS.CD	Net errors and omissions, adjusted (BoP, current US$)	International Trade